A Practitioner's Perspective on Simulation Validation

RPG Reference Document

8/15/01[1]

Table of Contents

---

[1] This document replaces the 11/30/00 version. It contains formatting and minor editorial changes.

This document corresponds to the web version of the VV&A RPG Reference Document of the same name and date. It has been modified to make it suitable for printing.

Table of Contents continued

What Is Simulation Validation and Why Is It Important?

Use of a model or simulation is a surrogate for experimentation with an actual system (existing or proposed), where experimentation with that system could be disruptive, not cost effective, or infeasible. If the model or simulation is unable to provide valid representations of the actual system, any conclusions derived from the model or simulation are likely to be erroneous and may result in poor decisions being made. Validation can be performed for **all** models and simulations, regardless of whether the corresponding real-world system exists in some form or will be built in the future. Validation should always be focused on the intended use.

A precise definition of validation is

> **Validation** is the **process** of determining the degree to which a model or simulation is an accurate representation of the real world from the perspective of the intended uses of the model or simulation.

The following are some general perspectives on validation:

- Conceptually, if a simulation is "valid," then it can be used to make decisions about the system similar to those that would be made if it were feasible and cost effective to experiment with the system itself.
- The ease or difficulty of the validation process depends on the complexity of the system being modeled and on whether a version of the system currently exists (see *Validating the Output from the Overall Simulation*).

> **Example 1**. A model of a neighborhood bank would be relatively easy to validate since it could be closely observed. However, a model of the effectiveness of a naval weapons system in the year 2025 would be virtually impossible to validate completely, since the location of the battle and the nature of the enemy weapons would be unknown. Also, it is often possible to collect data on an existing system that can be used for building and validating a model.

- A simulation of a complex system can only approximate the actual system, no matter how much time and money are spent on simulation construction. There is no such thing as absolute simulation validity, nor is it even desired. Indeed, a model or simulation is supposed to be an abstraction and simplification of reality. However, the most valid simulation is not necessarily the most cost effective. For example, increasing the validity of a simulation beyond a certain level might be quite expensive, since extensive data collection may be required, but might not lead to significantly better insight or decisions.

- A simulation should always be developed for a particular set of objectives. Indeed, a simulation that is valid for one set of objectives may not be for another set of objectives.
- The measures and acceptability criteria (e.g., measures of performance [MOPs])[1] used to validate a simulation should include those that the decision-maker will actually use for evaluating system configurations.
- Validation of a stand-alone simulation is a process that should be conducted in coordination with the development or modification effort. It is not something to be attempted after the simulation has already been developed (or modified) and then only if there is time and money remaining.
- A federation of models still has to be validated even if the models (federates) that compose it are believed to be valid.

Example 2. A military organization paid a consulting company $500,000 to perform a 6-month "simulation study." After the study was supposedly completed, a person from the client organization called the consulting company and asked, "Can you tell me in *5 minutes* on the phone how to validate our model?"

A model or simulation, its data, and its results have **credibility** if the decision-maker and other key project personnel accept them as "correct." Note that a credible simulation is not necessarily valid, and vice versa. The following factors help establish credibility for a model or simulation:

- decision-maker's understanding and agreement with the simulation's assumptions
- demonstration that the simulation has been validated and verified
- decision-maker's ownership of and involvement with the project
- reputation of the simulation developers

A model or simulation[2] that is **both** valid and credible is more likely to be formally accredited for use in a particular application.

The remainder of this document[3] is organized as follows:

[1] See the special topic on Measures for additional information.

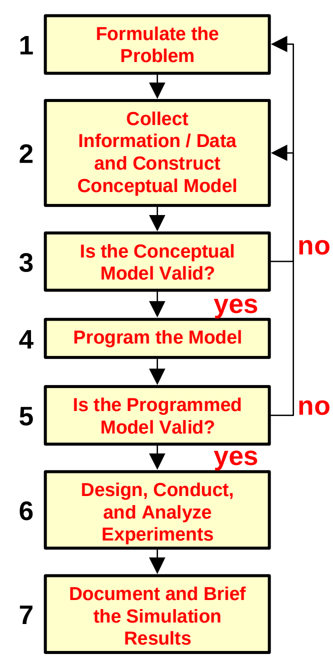
[2] In the remainder of this document, the term "simulation" will be used to mean either a model or a simulation, even though a simulation may be considered to be a particular kind of model [Law and Kelton, 1991, p. 4]. This word choice was made for pedagogical reasons and because most models used by the Department of Defense (DoD) are actually simulations.

[3] Many of the ideas and examples presented in this document are based on the chapter "Building Valid, Credible, and Appropriately Detailed Models" in Law and Kelton [1999]. Additional references are listed in the Reference section.

- Seven-Step Approach for Conducting a Successful Simulation Study presents a seven-step approach for conducting a successful simulation study
- Who Should Perform Validation? discusses who should perform validation of a simulation
- Techniques for Developing Valid and Credible Simulations presents techniques for developing a more valid and credible simulation
- Guidelines for Obtaining Good Data introduces guidelines for obtaining good simulation data
- Types of Simulation Development and Modification and Applicable Validation Techniques discusses different types of simulation development/modification and applicable techniques from *Techniques for Developing Valid and Credible Simulations*
- Summary presents a summary of important validation ideas

Seven-Step Approach for Conducting a Successful Simulation Study

The figure below depicts a seven-step approach for conducting a successful simulation study. Having a definitive approach for conducting a simulation study is critical to the study's success in general and to developing a valid simulation in particular. Each of the validation/credibility enhancement techniques defined in from *Techniques for Developing Valid and Credible Simulations* will be related to one or more of these steps.



Seven-Step Approach

Step 1. Formulate the problem

- The problem of interest is stated by the decision-maker (the User).
- A kickoff meeting for the simulation project is conducted with the project manager, simulation analysts (e.g., V&V Agent, Accreditation Agent), and subject matter experts (SMEs)[4] to discuss the following topics:
  – the overall objectives of the study
  – **specific** questions to be answered by the study

[4] See the special topic on Subject Matter Experts and VV&A for additional information on SMEs.

(without such specificity it is impossible to determine the appropriate level of simulation detail)

- the scope of the simulation (e.g., a single weapon system versus a battle with many weapons) and the level of simulation detail
- performance measures (e.g., metrics, acceptability criteria) that will be used to evaluate the efficacy of different system configurations
- system configurations
- required resources and time frame for the study

Step 2. Collect information/data and construct a conceptual model

- Collect information on the structure and operating procedures for the systems to be represented.
- Collect data[5] to specify simulation parameters and probability distributions (e.g., for the time to failure and the time to repair of an aircraft engine).
- The simulation assumptions, algorithms, and data summaries should be documented in a conceptual model[6] (assuming a conceptual model is in a written format).
- The level of simulation detail should depend on the following:
  - program objectives
  - performance measures of interest
  - data availability
  - credibility concerns
  - hardware constraints
  - opinions of SMEs[7]
  - time and money constraints
- Collect performance data from the existing systems (if any) (e.g., from a field test of a prototype system) to use in results validation in step 5

Step 3. Validate the conceptual model

- A structured walk-through of the conceptual model should be performed by a simulation analyst before an audience that includes the project manager, other

[5] See the reference document on M&S Data Concepts and Terms for additional information on data.
[6] See the special topic on Conceptual Model Development and Validation for additional information on conceptual model development.
[7] See the special topic on Subject Matter Experts and VV&A for additional information.

analysts, and SMEs is one method used to perform conceptual model validation.[8]

- If errors or omissions are discovered in the conceptual model, which is almost always the case, then the conceptual model should be updated before proceeding to step 4 (programming).

Step 4. Program the simulation

- Program the conceptual model in either a commercial simulation software product or in a general-purpose programming language (e.g., C or C++).
- Verify (debug) the computer code.

Step 5. Validate the programmed simulation

- If there is an existing (real-world) system, then compare simulation performance measures (acceptability criteria) with the analogous performance measures collected from the actual system (see step 2). This is called **results validation.** A modeled system should only represent those aspects of the real system that are needed for the specific application (e.g., the representation of a tank may need to include speed, armament, munitions, crew size, target acquisition capability but it may not need to include size, weight, color, fuel consumption, etc.). That combination of characteristics to be simulated in the representation is known as the **referent**[9] and it is against the referent that the simulation performance is measured.[10]
- Regardless of whether there is an existing system, the simulation analysts and SMEs should review the simulation results for reasonableness and to ensure the results are consistent with how they perceive the system should operate**.**
- Sensitivity analyses should be performed on the programmed representations to see which factors have the greatest effect on the performance measures and, thus, should be modeled most carefully.

Step 6. Design, make, and analyze simulation experiments

- For each system configuration of interest, tactical issues should be decided, such as run length and the number of independent simulation replications (multiple replications using different random numbers are generally required for stochastic models)

[8] See the special topic on Conceptual Model Development and Validation for information on conceptual model validation.
[9] Because a referent can be defined for nonexistent systems (e.g., futuristic weapon systems), results validation can be performed on a simulation even when actual data does not exist because the comparison is made to the referent.
[10] See the special topic on Validation for additional information.

- Results should be analyzed to determine if additional experiments are required.

Step 7. Document and brief the simulation results

- Documentation for the simulation (and the associated study) should include the conceptual model (critical for future reuse of the simulation), a detailed description of the computer program, and the results of the current study.
- The final briefing for the simulation study should include documentation and a discussion of the simulation development/validation process to promote simulation credibility.

Who Should Perform Validation?

Simulation validation (as compared to programming and verification) requires more analysis expertise than programming expertise. For example, in the case of discrete-event simulation, those developing and validating the simulation should be analysts with strong backgrounds in simulation methodology (e.g., validation techniques, selecting input probability distributions, design and analysis of simulation experiments), stochastic processes, probability theory, and statistics. Simulation development/modification and validation should be done hand-in-hand throughout the entire simulation life cycle. In fact, simulation validation does not end with the current simulation application but is continued with each new use of the simulation.[11]

Techniques for Developing Valid and Credible Simulations

Ten **practical** techniques for developing valid and credible simulations are discussed in the following sections. Each technique can be applied in one or more steps in the Seven-Step Approach, as shown in the table below.

Practical Techniques for Developing Valid, Credible Simulations	Applicable in Step Nos.
Formulating the Problem Precisely	1
Interviewing Subject Matter Experts	1, 2
Interacting with the Decision-Maker Regularly	1, 2, 3, 4, 5, 6, 7
Using Quantitative Techniques to Validate Components of the Simulation	2
Documenting the Conceptual Model	2
Performing a Structured Walk-through of the Conceptual Model	3

[11] Validation, like accreditation, is application specific: what is a valid representation of terrain in one application may not be valid in another.

Practical Techniques for Developing Valid, Credible Simulations	Applicable in Step Nos.
Performing Sensitivity Analyses to Determine Important Simulation Factors	5
Validating the Output from the Overall Simulation	5
Using Graphical Plots and Animations of the Simulation Output Data	5, 6, 7
Using Statistical Techniques for Comparing Simulation and System Output Data	5

Formulating the Problem Precisely

It is critical to formulate the problem of interest precisely. This formulation should include an overall statement of the problem to be solved, a list of the specific questions that the simulation is to answer, and the performance measures that will be used to evaluate the efficacy of particular system configurations. Without a definitive statement of the specific questions of interest, it is impossible to decide on the appropriate level of simulation detail. Performance measures must also be stated clearly since different measures may dictate different levels of simulation detail [see Law and Kelton, 1991, pp. 706-707 for more information].

When the decision-maker first initiates a simulation study, the exact problem to be solved is sometimes not precisely stated or even completely understood. Thus, as the study proceeds and is better understood, this new understanding should be communicated to the decision-maker who may reformulate the problem.

Interviewing Subject-Matter Experts

No one person can know all of the information necessary to build a simulation. Thus, it is necessary for the simulation analysts to talk to many different SMEs[12] to gain a complete understanding of the system or systems to be represented. Note that some of the information supplied by the SMEs will invariably be incorrect – if a particular part of the system is particularly important, then at least two SMEs should be queried. *Performing a Structured Walk-through of the Conceptual Model* discusses a technique that helps ensure that a simulation's assumptions are correct and complete – this technique is also useful for resolving differences of opinion among SMEs.

Interacting with the Decision-Maker Regularly

To develop a valid and credible simulation, the analyst should interact with the decision-maker and other members of the project team on a regular basis. This approach has the following key benefits:

- It helps ensure that the correct problem is solved in cases where

[12] See the special topic on Subject Matter Experts and VV&A for additional information.

- the exact nature of the problem may not be initially known
- the decision-maker may change the objectives during the course of the study
- the decision-maker may change during the study

- It helps maintain the decision-maker's interest and involvement in the study
- It helps make the simulation more credible because the decision-maker understands and agrees with the simulation's assumptions

> **Example 3.** A military analyst worked on a simulation project for several months without interacting with the general who requested it. At the final Pentagon briefing for the study, the general walked out after five minutes stating, "That's not the problem I'm interested in."

Using Quantitative Techniques to Validate Components of the Simulation

The analyst should use quantitative techniques whenever possible to test the validity of various components of the overall simulation. Examples of techniques that have been used for this purpose include.

- **Graphical plots and goodness-of-fit tests**. If one has fit a theoretical probability distribution (e.g., normal, exponential) to a set of observed data, then the adequacy of the representation can be assessed by using graphical plots and goodness-of-fit tests [Law and Kelton, 1991, chapter 6].
- **Data validation**. As discussed in *Guidelines for Obtaining Good Data*, it is important to use appropriate data in building a simulation; however, it is equally important to exercise care when structuring these data. For example, if several sets of data have been observed for the "same" random phenomenon, then the correctness of merging these data sets can be assessed by using the Kruskal-Wallis test of homogeneity of populations [Law and Kelton, 1991, pp. 409-410]. If the data sets appear to be homogeneous, they can be merged and the combined data set can be used for some purpose in the simulation.

> **Example 4**. Consider a manufacturing system for military aircraft, for which time-to-failure and time-to-repair data were collected for two "identical" machines made by the same vendor. However, the Kruskal-Wallis test showed that the two distributions were, in fact, different for the two machines. Thus, each machine was given its own time-to-failure and time-to-repair distributions in the simulation.

- **Factor analysis.** Rousseau and Bauer [1996] used factor analysis to identify a weakness in one of the mission-planning algorithms for the Airlift Flow Model, which is a large-scale transportation simulation that is primarily used for estimating strategic military-airlift performance.

Documenting the Conceptual Model

Communication errors often contribute to the invalid assumptions and inconsistencies in simulations. Documenting all simulation concepts, assumptions, objectives, requirements, algorithms, data summaries, etc. (i.e., developing a conceptual model) at the beginning of the simulation development process can lessen this problem. The conceptual model[13] of a simulation is the bridge between the simulation developer and the decision-maker. It denotes the developer's understanding of the problem objectives and requirements. Conceptual model documentation should be a major source of information about the simulation and should be readable by analysts, SMEs, programmers, and decision-makers. Some of the elements that should be included in conceptual model documentation are

- an overview section that discusses overall project goals, specific issues to be addressed by the simulation, and relevant performance measures
- diagrams showing the layout for the systems being represented
- detailed descriptions of each subsystem (preferably in bullet format for easy reading) and how they interact
- what simplifying assumptions (regarding representations and interactions) were made and why
- summaries of input data metadata[14]
- sources of important or controversial information

The conceptual model should contain enough detail so that it is a "blueprint" for creating the simulation computer program (in Step 4). Technical analyses and detailed rationales should be put in appendices to promote readability by decision-makers.

Performing a Structured Walk-through of the Conceptual Model

The developers of the conceptual model need to collect and synthesize information from many different sources. As a result, it is difficult to obtain a complete, consistent, and correct description of the overall system to be simulated. Conducting a **structured walk-through**[15] of the conceptual model before an audience of SMEs and decision-makers can minimize this problem. For example, using a projection device, the analyst reviews each item of the conceptual model with the review team to verify that all parts of the conceptual model are correct and at an appropriate level of detail. Conducted in

[13] See the special topic on Conceptual Model Development and Validation for additional information on conceptual models.
[14] See the reference document on M&S Data Concepts and Terms for additional information on metadata.
[15] See the reference document on V&V Techniques for additional information on walk-throughs.

this way, a structured walk-through can provide evidence that the developer has a correct understanding of the requirements of the problem and that the conceptual model is a valid foundation upon which to build the simulation.

The structured walk-through should ideally be held at a remote site (e.g., a hotel meeting room), so that the participants can give the meeting their full attention. Also, it should be held before programming begins in case major problems are uncovered. The conceptual model and documentation on the problem objectives and requirements should be sent to participants before the meeting so they have an opportunity to review them and prepare comments and recommendations. However, this should not be considered a substitute for the structured walk-through itself, since people may not the have the time or motivation to review the document carefully on their own. Furthermore, the interactions that take place at the actual meeting are invaluable.

It is highly unlikely that all aspects of the conceptual model will be found to be correct in a structured walk-through. Any errors or omissions should be corrected before programming begins (Step 4).

The above discussion assumes that the development of the simulation takes place all at the same time. However, for some incremental development paradigms[16] (e.g., spiral), there may have to be several structured walk-throughs (e.g., after each major stage of simulation development).

Performing Sensitivity Analyses to Determine Important Simulation Factors

An important technique for determining which simulation factors have a significant impact on the desired measures of performance is ***sensitivity analysis***. If a particular factor appears to be important, then it needs to be modeled carefully. The following are examples of factors that could be investigated by a sensitivity analysis:

- parameter values (see example 5)
- probability distribution selection
- entities moving through the simulated system
- level of detail for a subsystem
- data that are the most crucial to collect (a "coarse" model is used to identify the most important parts of the system)

[16] See the special topic on Paradigms for M&S Development for additional information on incremental paradigms.

Example 5. In a simulation study of a new weapons system, suppose that the value of a probability of kill is estimated to be 0.75 as a result of conversations with SMEs. The importance of getting the value of this probability "exactly" correct can be determined by running the simulation with 0.75 and, for example, by running it with each of the values 0.70 and 0.80. If the three simulation runs produce approximately the same results, then the output is not sensitive to the choice of the parameter *over the range 0.70 to 0.80*. Otherwise, a better specification of the probability is needed. (Strictly speaking, to determine the effect of the probability of kill on the simulation's results, several independent replications of the simulation should be made using different random numbers for each of the three cases.)

If one is trying to determine the sensitivity of the simulation output to changes in two or more factors of interest, then it is not correct, in general, to vary one factor at a time while setting the other factors at some arbitrary values. A more correct approach is to use statistical experimental design, which is discussed in Law and Kelton [1991, Chapter 12] and in Montgomery [1997]. The effect of each factor can be formally estimated and, if the number of factors is not too large, interactions between factors can also be detected.

Validating the Output from the Overall Simulation

The most definitive test of a simulation's validity is establishing that its output data closely resemble the output data that would be observed from the actual system. When it is not always possible to obtain data from the actual system (e.g., the system does not exist), then validation data must be obtained another way.

Using Data from Existing Systems

If a system similar to the proposed one now exists, then a simulation of the existing system is developed and its output data are compared to those from the existing system itself. If the two sets of data compare "closely," then the model of the **existing** system is considered "valid." (The accuracy required from the simulation will depend on its intended use and the utility function of the decision-maker.) The simulation is then modified so that it represents the proposed system. Greater commonality between existing and proposed systems leads to greater confidence in the simulation of the proposed system. There is no completely definitive approach for validating the simulation of the proposed system. If there were, then there might be no need for a simulation in the first place. If the above comparison is successful, then it has the additional benefit of providing credibility for the use of simulation. (As discussed in step 5, comparing simulation and system output data for the existing system is called results validation.)

Example 6. A U.S. Air Force test agency performed a simulation study for a wing of bombers using the Logistics Composite Model (LCOM). The goal of the study was to evaluate the effect of various proposed logistics policies on the availability of the bombers, i.e., the proportion of time that the bombers were available to fly missions. Data were available from the actual operations of the bomb wing over a 9-month period, and included both failure data for various aircraft components and a bomb-wing availability of 0.9. To validate the model, the Air Force first simulated the 9-month period with the existing logistics policy and obtained a model availability of 0.873, which is 3 percent different from the historical availability. This difference was considered acceptable because an availability of 0.873 would still allow enough bombers to be available for the Air Force to meet its mission requirements.

Example 7. The U.S. Army is thinking of purchasing a weapons system for which it is infeasible or too expensive to perform a complete set of evaluation tests. As an alternative, a simulation of the system is developed, and then a prototype of the actual system is field-tested on a military reservation for one or more specified scenarios. If the simulation and system output data compare closely for each of the specified scenarios, the "validated" simulation is used to evaluate the system for scenarios for which system field tests are not possible. Of course care must be taken in extrapolating beyond the scenarios for which the simulation was "validated."

Using Statistical Tests

A number of statistical tests (e.g., *t*, Mann-Whitney) have been suggested in the validation literature for comparing the output data from a stochastic simulation with those from the corresponding real-world system [Shannon, 1975, p. 208]. However, the comparison is not as simple as it might appear, since the output processes of almost all real-world systems and simulations are **non-stationary** (the distributions of the successive observations change over time) and **auto-correlated** (the observations in the process are correlated with each other). Thus, classical statistical tests based on independent, identically distributed (IID) observations are not **directly** applicable. Furthermore, it is questionable whether hypothesis tests, as compared with constructing confidence intervals for differences, are even the appropriate statistical approach. Since the simulation only approximates the actual system, a null hypothesis that the system and simulation are the "same" is clearly false. It is more useful to ask whether or not the differences between the system and the simulation are significant enough to affect any conclusions derived from the simulation. For a discussion of statistical procedures for comparing simulation and system output data, see *Statistical Techniques for Comparing Simulation and System Output Data*.

Consulting SMEs

Whether or not there is an existing system, analysts and SMEs[17] should review simulation output (numerical results, animations, etc.) for reasonableness. Face validation[18] is used to determine if simulation results are consistent with perceived

[17] See the special topic on Subject Matter Experts and VV&A for additional information.
[18] See the reference document on V&V Techniques for more information on face validation.

system behavior. However, care should be taken in performing this exercise, since if one knew exactly what output to expect, then there would be no need for a simulation.

Example 8. Face validation was used in the development of a simulation of the U.S. Air Force manpower and personnel system. (This simulation was designed to provide Air Force policy analysts with a system-wide view of the effects of various proposed personnel policies.) The simulation was run under the baseline personnel policy, and the results were shown to Air Force analysts and decision-makers, who subsequently identified some discrepancies between the simulation and perceived system behavior. This information was used to improve the simulation, and after several additional evaluations and improvements, a simulation was obtained that appeared to approximate current Air Force policy closely. This exercise improved not only the validity of the simulation, but also its credibility.

A **Turing test** [see Turing, 1950 and Carson, 1986] can be used to compare output data from the simulation to those from the real system. People knowledgeable about the system (e.g., SMEs or decision-makers) are asked to examine one or more sets of system data as well as one or more sets of simulation data without knowing which sets are which. Each data set should be presented on a separate piece of paper using exactly the same format. If the SMEs can differentiate between the system and simulation data, their explanation of how they were able to do so is used to improve the simulation.

Example 9. Schruben [1980] reports the use of a Turing test in a simulation study of an automobile component factory. Data from the factory and from the simulation were put on time-study forms and reviewed at a meeting by three managers, three industrial engineers, and two factory workers. The inability of these people to agree on which data were real and which were simulated led to immediate acceptance of the simulation.

Another technique used to validate a simulation is to compare its results with those from a simulation known to be "valid" for the application of interest.

Example 10. A defense supply center was building a new simulation called the Performance and Requirements Impact Simulation to replace an existing simulation. One of the purposes of both simulations is to decide when to order and how much to order for each stock number.

To validate the **old model**, the total dollar amount of all orders placed by the model for fiscal year 1996 was compared to the total dollar amount for the actual system for the same time period. Since these dollar amounts differed by less than 3 percent, there was a fair amount of confidence in the validity of the old model.

In order to validate the **new model**, the two simulations were used to predict the total dollar amount of all orders for fiscal year 1998 and the results differed by less than 6 percent. Thus, there was reasonable confidence in the validity of the new simulation.

Prospective Validation

Up to now the discussion has focused on validating a simulation relative to past or present system output data; however, a perhaps more definitive test of a simulation is to establish its ability to predict **future** system behavior. Since simulations often evolve over time and are used for multiple applications, there is often an opportunity for such **prospective** validation. For example, if a simulation is used to decide which version of a proposed system to build, then after the system has been built and sufficient time has elapsed for output data to be collected, these data can be compared with the predictions of the simulation. If there is reasonable agreement, confidence in the "validity" of the simulation increases. However, discrepancies between the two data sets should be used to update the simulation. Regardless of the accuracy of the past predictions, a simulation should be carefully scrutinized before each new application, since a change in purpose or the passage of time may have invalidated some aspect of the existing simulation. This points out the need for good simulation documentation (e.g., a conceptual model).

Using Graphical Plots and Animations of the Simulation Output Data

Graphical plots (static or dynamic) and animations (dynamic) are useful for showing that a simulation is **not** valid as well as for promoting simulation credibility. The following are some examples of graphical plots:

- histogram (a graphical estimate of the underlying probability density or mass function)
- correlation plot (shows if the output data are auto-correlated)
- time plot (one or more simulation variables are plotted over the length of the simulation run to show the **long-run** dynamic behavior of the system)
- bar charts and pie charts

An animation, which shows the **short-term** dynamic behavior of a system, is useful for communicating the essence of a simulation to decision-makers and other people who do not understand or care about the technical details of the simulation. It is a great way to enhance the credibility of a simulation. Animations are also useful for verification of the simulation computer program, for suggesting improved operational procedures, and for training.

Using Statistical Techniques for Comparing Simulation and System Output Data

In this section statistical procedures are presented that might be useful for comparing simulation and system output data as discussed in *Validating the Output from the Overall Simulation*.

Suppose that R_1, R_2,…, R_k are observations from a real-world system and that M_1, M_2,…, M_l are output data from a corresponding simulation (see example 11). These data sets should be compared in some way to determine whether the simulation is an accurate representation of the real-world system. However, most classical statistical approaches such as confidence intervals and hypothesis tests assume that the real-world data and the simulation data are each IID data sets, which is generally not the case. Thus, these classical statistical approaches are not **directly** applicable to our comparison problem.

> **Example 11.** Consider a military communications system where the data of interest are the end-to-end delays of successively completed messages. These data are not independent for the actual system (or for a corresponding simulation). For example, if the system is busy at a particular point in time, then all of the messages being processed will tend to have large delays (i.e., the delays are **positively correlated**).

The inspection, confidence interval, and time-series approaches for comparing simulation and system output data are discussed in the next three sections. A much more comprehensive discussion of these topics can be found in Law and Kelton (1991, Section 5.6).

Inspection Approach

Simulation practitioners who compare simulation and system output data compute one or more statistics from the real-world observations and corresponding statistics from the simulation output data, and then compare the two sets of statistics without the use of a formal statistical procedure. Examples of statistics that might be used for this purpose are the sample mean, the sample variance [see Law and Kelton, 1991, Section 4.4, for a discussion of the danger in using the sample variance from auto-correlated data)], the sample correlation function, and *histograms*. (*Histogram* is italicized because histograms are usually derived from IID data.) Sargent [1996b] discusses the use of graphical plots in more detail, which a good idea when applicable. The difficulty with this inspection approach,[19] which was used in example 6, is that each statistic is essentially a sample of size 1 from some underlying population, making this idea vulnerable to the inherent randomness of the observations from both the real system and the simulation.

> **Example 12**. For the communications network of example 11, an application of the inspection approach might compare the sample mean of the end-to-end delays for the simulation with the sample mean of the end-to-end delays for the system.

Confidence Interval Approach

[19] See the reference document on V&V Techniques for additional information on inspection.

This is a more reliable approach for comparing a simulation with the corresponding system when it is possible to collect several independent sets of data from the system. This confidence interval approach is discussed in Appendix A.

Time-Series Approaches

There have been several time-series approaches suggested for comparing simulation output data with system output data. [A **time series** is a finite realization of a stochastic process. For example, the end-to-end delays E_1, E_2, …, E_{100} from the communications system (or the corresponding simulation) of the example in a confidence interval approach described in Appendix A form a time series.] These approaches require only **one** set of each type of output data and may also yield information on the auto-correlation structures of the two output processes. However, these approaches, which are based on spectral analysis, parametric time-series models, or standardized time series, all make certain assumptions about the simulation and system output data that may not be satisfied in practice. They also require a much higher level of mathematical sophistication to apply than the *Confidence Interval Approach*.

Guidelines for Obtaining Good Data

A simulation is only valid for a particular application if its logic is correct **and** if it uses appropriate data[20]. This section provides some suggestions on how to obtain good data.

Two Basic Principles

If a system similar to the one of interest exists, then data should be obtained from it for use in building the simulation. These data may be available from historical records or may have to be collected during field tests. Since the people who provide the data are generally different from the simulation analysts, the analysts need to

- make sure that the required data are specified precisely in terms of type, format, amount, conditions under which it should be collected, why it is needed, etc. (i.e., data quality)[21] to the people providing the data
- understand the process that produced the data, rather than treating the observations as just abstract numbers

[20] See the reference document on M&S Data Concepts and Terms for additional information on M&S data.
[21] See the Data Quality Templates for more information on specifying data quality for simulation use.

Example 13. Suppose that data are available on the time to perform some task (e.g., repair an aircraft engine), but a few observations are significantly larger than the rest (called *outliers*). Without a good understanding of the underlying process, it is impossible to know whether these large observations are the result of measuring or recording errors, or are just legitimate values that occur with small probability.

Common Difficulties

The following are five potential difficulties with data. Data may

- not be representative of what needs to be simulated

Example 14. The data that have been collected during a military field test may not be representative of actual combat due to differences in troop behavior and unrealistic battlefield conditions (e.g., lack of smoke).

- not be of the appropriate type or format
- contain measuring, recording, or rounding errors

Example 15. Data representing the time needed to perform some task are sometimes rounded to the closest 5 or 10 minutes. This may make it difficult to fit a continuous theoretical probability distribution to the data, since the data are now discrete.

- be "biased" because of self-interest
- employ inconsistent units

Example 16. The U.S. Transportation Command transports military cargo by air, land, and sea. Sometimes there is confusion in building simulations because the U.S. Air Force and the U.S. Army use short tons (2000 pounds) while the U.S. Navy uses long tons (2200 pounds).

Data V&V done in conjunction with simulation V&V can detect these and other problems and help ensure the data selected are appropriate for use in the simulation.[22]

Types of Simulation Development and Modification and Applicable Validation Techniques

[22] See the special topics on Data V&V for New Simulations, Data V&V for Legacy Simulations, and Data V&V for Federations for more information.

When analysts decide to study a problem by simulation, they must decide which of the following options is the most appropriate:

- develop a completely new simulation
- modify the logic of an existing or legacy simulation[23]
- use a legacy simulation with new data (the simulation logic is not changed)

The next three sections discuss which of the validation/credibility enhancement techniques in *Techniques for Developing Valid and Credible Simulations* are applicable in each of the above simulation development/modification situations.

New Simulation Development

If an analyst is developing a new simulation, then all of the techniques discussed in *Techniques for Developing Valid and Credible Simulations* are potentially applicable to simulation development and validation. Guidelines for deciding on an appropriate level of simulation detail are discussed in Law and Kelton [1999, Chapter 5].

Modifying Legacy Simulation Logic

All of the techniques discussed in *Techniques for Developing Valid and Credible Simulations* are potentially applicable when modifying a legacy simulation; however, the techniques in *Documenting the Conceptual Model* are especially useful. A lack of good documentation (e.g., problem statement, requirement definitions, conceptual model, design products, data descriptions, documented code) makes modifying a legacy simulation difficult. Random comments embedded in the computer program are not sufficient. Even if an analyst had the time and motivation to try to understand the existing program, there would still be certain implicit simulation assumptions that would be missed.

If no conceptual model exists for the legacy simulation, one should be constructed for the modified simulation. Although this conceptual model should focus on providing information (e.g., assumptions, entities, scenarios, algorithms) about the modified portion of the simulation, it should include as much information as possible about the part of the legacy simulation that will remain after the modification

Using a Legacy Simulation with New Data

A legacy simulation in which the logic doesn't change (i.e., no major modifications are made) but that will be used with new data (e.g., new munitions, new scenarios, different force structure) still needs validation. When data values change, there is no guarantee

[23] The DoD Modeling and Simulation Resource Repository (MSSR) is one source of information on legacy simulations and models within the DoD.

that the simulation will still produce valid results. The new data values could be outside the range of validity for which the simulation was originally developed.

All the techniques discussed in the *Techniques for Developing Valid and Credible Simulations* sections[24] potentially apply ; however,

- *Interacting with the Decision-maker Regularly*
- *Documenting the Conceptual Model*
- *Performing a Structured Walk-through of the Conceptual Model*

are most beneficial because they correspond to the collection, documentation, and (formal) review, respectively, of the data.[25]

> **Example 17.** LCOM is an example of a legacy simulation that is typically used with new data, without changes to its logic. Examples of data that might need to be specified for a particular application are the number of bombers in the wing, the frequency of missions (e.g., every two to three days), and the mission duration.

Summary

All simulations need to be validated or any decisions made with them may be erroneous. The amount of validation that is done on a particular simulation for a given application is dependent on

- the risk[26] associated with using an inaccurate simulation to make decisions
- the cost of collecting additional evidence (e.g., by performing additional validation tasks, by conducting more field tests) about the validity of the simulation

The following table summarizes the basic procedure for developing a valid and credible simulation.

Developing a Valid and Credible Simulation
• Formulate the problem precisely
• Interview appropriate SMEs
• Interact with the decision-maker on a regular basis -- to ensure that the correct problem is being solved and to promote simulation credibility

[24] Except those described in *Using Quantitative Techniques to Validate Components of the Simulation* and *Performing Sensitivity Analyses to Determine Important Simulation Factors*
[25] See the special topic on Data V&V for Legacy Simulations for more information.
[26] See the special topic on Risk and Its Impact on VV&A.

Developing a Valid and Credible Simulation
• Validate components of the simulation – using quantitative techniques
• Document the conceptual model – critical for current and future applications of the simulation
• Perform a structured walk-through of the conceptual model – for a nonexistent system, this may be the single most-important validation technique
• Perform sensitivity analyses to determine important simulation factors and risks
• Validate simulation results – analyzing simulation output data using various techniques

References

Balci, O., "Verification, Validation and Testing," in *The Handbook of Simulation*, J. Banks, ed., Chapter 10, John Wiley, New York (1998).

Banks, J., J. S. Carson, and B. L. Nelson, *Discrete-Event System Simulation,* Second Edition, Prentice-Hall, Upper Saddle River, NJ (1996).

Carson J. S., "Convincing Users of Model's Validity Is Challenging Aspect of Modeler's Job," *Ind. Eng., 18:* 74-85 (June 1986).

Fossett, F. A., D. Harrison, H. Weintrob, and S. I. Gass, "An Assessment Procedure for Simulation Models: A Case Study," *Operations Res*., 39*:* 710-723 (1991).

Gass, S. I., "Decision-Aiding Models: Validation, Assessment, and Related Issues in Policy Analysis," *Operations Res., 31:* 603-631 (1983).

Gass, S. I. and B. W. Thompson, "Guidelines for Model Evaluation: An Abridged Version of the U.S. General Accounting Office Exposure Draft," *Operations Res., 28:* 431-439 (1980).

Knepell, P. L. and D. C. Arangno, "*Simulation Validation: A Confidence Assessment Methodology*," IEEE Computer Society Press, Los Alamitos, CA (1993).

Law, A. M. and W. D. Kelton, *Simulation Modeling and Analysis,* Second Edition, McGraw-Hill, New York (1991).

Law, A. M. and W. D. Kelton, *Simulation Modeling and Analysis,* Third Edition, McGraw-Hill, New York (December 1999).

Montgomery, D. C., *Design and Analysis of Experiments*, 4th Edition, John Wiley, New York (1997).

Pace, D. K., "Verification, Validation, and Accreditation (VV&A)," in *Applied Modeling and Simulation: An Integrated Approach to Development and Operation*, D. J. Cloud and L. B. Rainey, eds., pp. 369-410, McGraw-Hill, New York (1998).

Rousseau, G. G. and K. W. Bauer, "Sensitivity Analysis of a Large Scale Transportation Simulation Using Design of Experiments and Factor Analysis," *Proc. 1996 Winter Simulation Conference*, San Diego, pp. 1426-1432 (1996).

Sargent, R. G., "Verifying and Validating Simulation Models," *Proc. 1996 Winter Simulation Conference,* San Diego, pp. 55-64 (1996a).

Sargent, R. G., "Some Subjective Validation Methods Using Graphical Displays of Data," *Proc. 1996 Winter Simulation Conference*, San Diego, pp. 345-351 (1996b).

Schruben, L. W., "Establishing the Credibility of Simulations," *Simulation, 34:*101-105 (1980).

Shannon, R. E., *Systems Simulation: The Art and Science*, Prentice-Hall, Englewood Cliffs, NJ (1975).

Turing, A. M., "Computing Machinery and Intelligence," *Mind, 59:* 433-460, (1950).

RPG References in This Document

select menu: *RPG Reference Documents*, select item: "M&S Data Concepts and Terms"

select menu: *RPG Reference Documents*, select item: "V&V Techniques"

select menu: *RPG Special Topics*, select item: "Data V&V for Federations"

select menu: *RPG Special Topics*, select item: "Data V&V for Legacy Simulations"

select menu: *RPG Special Topics*, select item: "Data V&V for New Simulations"

select menu: *RPG Special Topics*, select item: "Conceptual Model Development and Validation"

select menu: *RPG Special Topics*, select item: "Measures"

select menu: *RPG Special Topics*, select item: "Paradigms for M&S Development"

select menu: *RPG Special Topics*, select item: "Risk and Its Impact on VV&A"

select menu: *RPG Special Topics*, select item: "Subject Matter Experts and VV&A"

select menu: *RPG Special Topics*, select item: "Validation"

select menu: *RPG Templates*, select item: "Data Quality Templates"

In the web-based version of this document, the appendix below appears as a hot link in the section on *Using Statistical Techniques for Comparing Simulation and System Output Data (Confidence Interval Approach).*

Appendix A: Confidence Interval Approach Based on Independent Data

Collect n independent sets of data from the system and n independent sets of data from the simulation. (If n independent replications of the simulation are made using different random numbers, then the n resulting sets of simulation output data are independent of each other.)

Let X_j be the sample mean of the observations in the jth set of system data, and let Y_j be the sample mean of the observations in the jth set of simulation data. The X_j's are IID random variables (assuming that the n sets of system data are homogeneous) with mean $\mu_X = E(X_j)$, and the Y_j's are IID random variables with mean $\mu_Y = E(Y_j)$. [If Z is a random variable, then $E(Z)$ is its expected value or mean.] Compare the simulation with the system by constructing a confidence interval for $\zeta = \mu_X - \mu_Y$.

Let $Z_j = X_j - Y_j$ for $j = 1, 2, \ldots, n$.

Also let the sample mean, $\bar{Z}(n)$, and the sample variance, $S^2(n)$, be defined as follows:

$$\bar{Z}(n) = \sum_{j=1}^{n} Z_j / n$$

and

$$S^2(n) = \sum_{j=1}^{n} [Z_j - \bar{Z}(n)]^2 / (n-1)$$

Then, an approximate $100(1 - \alpha)$ percent confidence interval ($0 < \alpha < 1$) for $\zeta = \mu_X - \mu_Y$ is given by

$$\bar{Z}(n) \pm t_{n-1,1-\alpha/2} \sqrt{S^2(n)/n}$$

where $t_{n-1, 1-\alpha/2}$ is the upper $1 - \alpha/2$ critical value[1] for a t distribution with $n-1$ degrees of freedom.

If the confidence interval is centered near 0 and has a short length, then this suggests that the simulation mean μ_Y is close to the system mean μ_X and that the simulation is "valid."

[1] These critical values are given in many statistics books and in Law and Kelton [1991, p. 738]. Also see pages 286-290 and 586-591 in Law and Kelton for a general discussion of confidence intervals and their proper interpretation.

Example. Suppose a prototype military communications system is field-tested on a military reservation for a specified scenario, and that the field test is repeated $n = 5$ times. Let X_j be the sample mean of the end-to-end delays for the jth trial of the field test. Suppose that a simulation is constructed and that $n = 5$ independent replications of the model are made. Let Y_j be the sample mean of the end-to-end delays for the jth replication of the model. Suppose further that a 95 percent confidence interval is constructed for $\zeta = \mu_X - \mu_Y$ and [–0.4, 0.6] (the units are seconds) are obtained. Thus, there is 95 percent confidence that the system mean differs from the model mean by between –0.4 second and 0.6 second for the scenario tested. Whether this is an acceptable difference depends on the issues that the model will be used to address.

§ § § § § § §